VOLUME 24 · NUMBER 1 · JANUARY 1 2006

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Comparison of CA-125 and Standard Definitions of Progression of Ovarian Cancer in the Intergroup Trial of Cisplatin and Paclitaxel Versus Cisplatin and Cyclophosphamide

Gordon J.S. Rustin, Petra Timmers, Ann Nelstrop, Gavin Shreeves, Soeren M. Bentzen, Benoit Baron, Martine J. Piccart, Kamma Bertelsen, Gavin Stuart, James Cassidy, and Elizabeth Eisenhauer

From the Department of Medical Oncology, Mount Vernon Cancer Centre, Northwood, Middlesex; Gray Cancer Institute, Mount Vernon Hospital, Northwood; Scottish Group Data Centre, Beatson Oncology Centre, Western Infirmary Glasgow, United Kingdom; European Organisation for Research and Treatment of Cancer (EORTC) –Gynecological Cancer Cooperative Group, EORTC Data Centre, Brussels, Belgium; Nordic Gynecological Cancer Study Group Centre, Odense University Hospital, Denmark; and National Cancer Institute of Canada, Clinical Trials Group Data Centre, Queen's University, Kingston, Ontario, Canada.

Submitted February 2, 2005; accepted July 20, 2005.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Address reprint requests to Gordon J.S. Rustin, MD, Department of Medical Oncology, Mount Vernon Cancer Centre, Northwood, Middlesex, HA62RN, United Kingdom. e-mail: grustin@nhs.net.

0732-183X/06/2401-45/$20.00

DOI: 10.1200/JCO.2005.01.2757

ABSTRACT

Purpose
A definition for progression of ovarian cancer has been proposed based on either a confirmed doubling of CA-125 levels from the upper limit of normal or from the nadir level if levels are persistently elevated. Retrospectively, we determined whether the use of this CA-125 definition in a randomized trial would have shown the same magnitude of difference between the treatment arms as was shown when the standard progression definition was used.

Patients and Methods
A retrospective analysis was performed on 680 patients in the Taxol Intergroup Trial with advanced epithelial ovarian carcinoma, of whom 628 were assessable according to CA-125. The date of progression according to clinical or radiologic criteria was compared with the date of progression according to CA-125.

Results
Of the 628 patients assessable for both definitions, 556 clinical or radiologic progressions were determined compared with 389 according to the CA-125 definition. There was a highly significant difference in the hazard of progression between the paclitaxel and cisplatin arm (TP) compared with the cyclophosphamide and cisplatin arm (CP) when either standard or CA-125 criteria were used to define progression (standard, $P = .002$; CA-125, $P = .011$). The hazard ratio of TP/CP over time was similar when comparing the different methods of defining progression.

Conclusion
The results of this analysis show that the magnitude of the therapeutic benefit was similar whether CA-125 or standard criteria were used to define progression.

J Clin Oncol 24:45-51. © 2006 by American Society of Clinical Oncology

INTRODUCTION

An increasing CA-125 has been shown to predate clinical or scan evidence of relapse in approximately 70% of patients with ovarian cancer by a median of 4 months.[1-3] It is increasingly accepted as an early indicator of disease relapse. In clinical trials of first- and second-line treatment for ovarian cancer in which progression-free survival (PFS) is a major end point, it is well known that many investigators will instigate second-line treatment because of an increase in serum CA-125 levels before clinical or radiologic signs of progression.[4] This then causes great confusion when determining the date of disease progression, with wide variability among different trial groups regarding handling of these data. Some groups include these patients in the progressive disease (PD) population at the date new therapy is started, other groups censor these patients at that date, and others ignore the new treatment before the documented date of clinical progression altogether. The Gynecologic Cancer Intergroup (GCIG) therefore proposed that a precise definition of progression be used as a secondary end point in clinical trials.[5]

Several definitions of progression according to CA-125 have been proposed. An increase of 50%, 100%, or just to above the normal range have all been shown to be predictive of relapse.[2,6-8] Rustin et al[9] produced and validated a definition that was based on a serial increase of 25% for four samples, 50% for three samples, or levels persistently elevated more than 100 U/mL, which required a computer program to maintain accuracy.

Two simpler definitions have now been produced. The first was developed after 255 patients were studied in the North Thames Ovary Trial of five versus eight courses of chemotherapy.[3] It was found that a confirmed increase of serum CA-125 to more than twice the upper limit of normal during follow-up after first-line chemotherapy predicted tumor relapse with a sensitivity of 84% and a false-positive rate of less than 2%. The second definition was developed after 88 patients were studied whose CA-125 levels remained persistently elevated during and/or after first-line chemotherapy. In this group a confirmed doubling of CA-125 from its nadir level predicted progression with a sensitivity of 94% and almost 100% specificity.[10]

The GCIG has produced a definition based on these last two definitions that also incorporates the Response Evaluation Criteria in Solid Tumors Group (RECIST) progression criteria[11] (Table 1). Although the GCIG definition is being incorporated into many protocols, it requires validation before being accepted as standard. This study attempts to help validate the GCIG definition by comparing the date of progression determined retrospectively using the CA-125 criteria with the date of progression that was defined prospectively using standard criteria in a large randomized trial.[12] We were particularly interested to know whether the CA-125 criteria can distinguish differences in treatment effect between trial arms with reasonable statistical power when an important difference exists by use of the standard definition; whether estimates of relative treatment effects given by CA-125 criteria are consistent with those given by standard criteria; and whether significant differences between trial arms can be detected earlier by use of CA-125 criteria than by use of standard criteria.

PATIENTS AND METHODS

This ovarian cancer Intergroup trial referred to here as the Taxol Intergroup Trial was led by the European Organisation for Research and Treatment of Cancer Gynecologic Cancer Group with participation by the National Cancer Institute of Canada, Nordic Gynecological Cancer Study Group Centre and Scottish Groups. It was a phase III study of patients with International Federation of Gynecology and Obstetrics stage IIb to IV ovarian cancer comparing the combination of cyclophosphamide and cisplatin (CP) with the combination of paclitaxel and cisplatin (TP), with full details published previously.[12] The trial was approved by the institutional review board of each contributing center. After optimal or suboptimal surgical debulking of the tumor, random assignment took place between CP and TP adjuvant chemotherapy. Patients received up to nine courses of cisplatin 75 mg/m^2 and cyclophosphamide 750 mg/m^2, or paclitaxel 175 mg/m^2 during 3 hours and cisplatin 75 mg/m^2. The protocol indicated that samples should be sent for CA-125 measurements on the first day of each chemotherapy cycle, every 3 months for the first 2 years of follow-up, and then once every 6 months. CA-125 was measured by one of the commercially available immunoradiometric assays (Centocor, Horsham, PA or Abbott Laboratories, Abbott Park, IL) and centers were presumed to have used the same assay method for an individual patient throughout the study period.

Definition of Study End Points

The date of PD according to standard criteria was defined as the date of unequivocal increase of at least 25% in the sum of the products of the perpendicular diameters of the measured lesions, or the appearance of new lesions. This definition of progression differs from the WHO definition in the use of the sum of the products of individual lesions, and also differs from the RECIST definition in Table 1. PFS was defined as the interval between the date of random assignment and the date of progression of the disease, death, or start of a new therapy without evidence of progression—whichever occurred first. The date of progression and definition of progression according to CA-125 were calculated according to the criteria listed in Table 1.

Two important limitations to this retrospective study should be highlighted. First, the diagnostic activity with respect to CA-125 measurements was not optimal. There are good reasons to believe that if CA-125 measurements had been required as an end point in the original protocol, CA-125 sampling would have been more complete. In addition, because the Taxol Intergroup Trial was designed long before the GCIG criteria were produced, many patients did not have a second sample to confirm biochemical progression. A confirmatory sample therefore was not required to define CA-125 progression in this analysis. An upper limit of normal of 23 U/mL was used throughout. The second caveat is that in comparing the various definitions of progression, we have to assume that the clinical and radiologic assessment is always correct. In other words, we are forced to use the terms false-positive and false-negative CA-125 predictions in relation to the status determined using standard response criteria. This is a built-in dichotomy: a false-positive CA-125, in principle, could be a false-negative prediction using standard criteria. Prolonged observation of the patient, in principle, could resolve this ambiguity. However, this generally was not possible in the present analysis and therefore the point of view taken in the following discussion is that the standard response evaluation is a gold standard against which our alternative definitions should be compared. The combination of the standard and CA-125 criteria assumed the date of progression to be the earliest date defined by either method.

Methods of Analysis

The analyses of PFS were based on intent-to-treat policy. The survival curves were estimated according to the Kaplan-Meier technique.[13] Differences in the time-to-event end points were compared with the use of a two-sided

Table 1. GCIG Definition of Progression of Ovarian Cancer According to Both Clinical and CA-125 Criteria

Criteria	Group A	Group B	Group C
Measurable/ assessable disease	Compared with baseline (or lowest sum while on study if less than baseline), a 20% increase in sum of longest diameters (RECIST definition) or Any new lesions (measurable or non-measurable) Date of PD Date of documentation of increase or new lesions		
CA-125	CA-125 elevated pre-treatment but later normalizes	CA-125 elevated pretreatment and does not normalize	CA-125 in normal range pretreatment
	CA-125 $\geq$ 2 $\times$ ULN documented on two occasions*	CA-125 $\geq$ 2 $\times$ nadir value on two occasions*	As for group A
	Date PD: first date of the CA-125 elevation to $\geq$ 2 $\times$ ULN	Date PD: first date of the CA-125 elevation to $\geq$ 2 $\times$ nadir value	

NOTE. Modified from Vergote et al.[5] Patient groups A, B, and C are defined according to CA-125 behavior during first-line therapy. A patient may be declared to have PD on the basis of either the objective disease or the CA-125 criteria. The date of PD will be the date of the earlier of the two events if both are documented.
Abbreviations: GCIG, Gynecologic Cancer Intergroup; RECIST, Response Evaluation Criteria in Solid Tumors Group; ULN, upper limit of normal; PD, progressive disease.
*Repeat CA-125 any time, but normally not less than 1 week after the first elevated CA-125 level. CA-125 levels sampled after administration of mouse antibodies or within 4 weeks after surgery or paracentesis should not be taken into account.

unstratified log-rank test.[14] We have also carried out analyses of the hazard of progression over time using the discrete time version of the proportional hazards model described by Aitkin et al.[15] This model is more informative than overall log-rank tests for survival; it can be used to assess whether the risk of progression is concentrated in a particular time window and can also reveal any interactions between elapsed time and trial arm.

We have also analyzed the distribution of lead and lag times between CA-125 doubling and clinical progression to provide an indication of any advantages in earlier diagnosis of progression, which clinicians could obtain by measuring CA-125 levels. We define a lead time as occurring when progression according to the CA-125 criteria occurs at or before the date of clinical progression. We define a lag time as occurring when CA-125 progression is observed after clinical progression.

Censoring and Data Inclusion

All patients registered were analyzed according to standard criteria but the duration of follow-up is longer than in the published report of the Taxol Intergroup Trial.[12] Only patients with at least two CA-125 samples were considered assessable according to CA-125. Although all patients are included in the Kaplan and Meier curves, we only used those assessable according to CA-125 in our hazard analysis to ensure that we were comparing groups that are equal in all respects except for the method used to assess progression. Patients were censored according to the CA-125 progression definition if their CA-125 levels had not doubled by the date when their last marker levels were measured. They were censored for clinical progression if they showed no signs of clinical progression by their last follow-up date. Because some patients' markers were not measured throughout their follow-up, their last date for CA-125 follow-up was not necessarily the same as their last follow-up date for clinical progression.

For the analysis of the lead/lag times between CA-125–defined progression and clinical progression, patients without either a known date of CA-125 progression or a known date of clinical progression cannot be included, given that they provide no information about whether they might eventually have shown a lag or a lead time. However, patients who have shown clinical progression without CA-125 doubling by the end of follow-up can be included because they have known but censored lag times.

Analysis of Hazard

We use the term hazard to mean the risk of experiencing progression during a given time interval. We categorized times from first treatment to progression or censoring into intervals 90 days wide, which allowed analyses to be performed using the discrete time proportional hazards model. In each analysis we have assessed the significance of trial arm, time interval, and potential interactions between time and trial arm using the GLIM statistical package.[15] Statistical significance is assessed in GLIM by general linear modeling methods in which terms are removed from a maximal model until additional removals lead to significant ($P < .05$) increases in deviance (G), leaving a minimal adequate model in which all remaining terms are significant. We report significance levels of terms as obtained on their stepwise removal in order of increasing significance. For additional details of the model and its analysis using GLIM see Aitkin et al.[15]

RESULTS

There were 680 patients entered into the Taxol Intergroup Trial, of whom clinical and or radiologic progression has been documented in 556. At the time of the last follow-up available for this study, which differs from the date used in the other publications on the trial, deaths had occurred in 435 patients and 245 were still living, of whom 124 had no evidence of progression and 121 were alive with progression. The median follow-up period was 897 days. There were 628 assessable according to CA-125 (Table 2). Among the patients who experienced progression according to CA-125, 250 had CA-125 levels that were initially increased but decreased to or below the upper limit of normal (group A, Table 1). There were 129 whose nadir level of CA-125 was above the upper limit of normal (group B) and there were 10 whose CA-125 values were $\leq$ 23 U/mL at the start of treatment (group C). The CA-125 nadir was $\leq$ 23 U/mL in 466 patients, between 24 and 99 U/mL in 121 patients, and $\geq$ 100 U/mL in 93 patients.

Accuracy of Doubling of CA-125

The accuracy of the detection of progression comparing clinical and CA-125 methods is shown in Table 3. There were 111 of 239 (46%) patients who were predicted by CA-125 not to have progressed and who had no documented clinical progression (ie, these are regarded here as true negative results). The remaining 128 of 239 patients (54%) had clinical progression but had not experienced progression according to the CA-125 criteria at the last available assessment. However, these are not necessarily false-negative predictions because a proportion of these patients would likely have experienced progression if prolonged observation of CA-125 had been available. The Kaplan-Meier estimate of the proportion of patients without CA-125 progression after clinical progression at 1 and 2 years were estimated at 28.1% $\pm$ 3.7% and 20.2% $\pm$ 3.7%, respectively. This corresponds to overall false-negative rates of 12.1% (95% confidence limits [CL] = 9.0%, 15.2%) and 8.7% (95% CL = 5.6%, 11.8%) in the group of 456 patients included in the lead/lag time analysis (see Lead and Lag Time of CA-125–Defined PD).

Among the assessable patients, only six (six of 389; 1.5%) had a false-positive prediction of progression by CA-125 corresponding to a false-positive rate among the assessable patients of 1.3% (95% CI, 0.5% to 2.8%). Of these, three have had only a short follow-up time since the date of CA-125 progression, and in two patients it was just a single (possibly rogue) sample that predicted progression; there is only one unexplained false-positive prediction. The specificity of CA-125 defined progression was 95%.

Table 2. Patient Population for Analyses

Characteristic	Arm 1	Arm 2	Total
All patients in dataset (Taxol Intergroup Trial)	338	342	680
Nonassessable for CA-125 progression criteria	30	22	52
Total No. in hazard analyses	308	320	628
With neither CA-125 doubling nor clinical progression dates	48	63	111
Assessable for lead/lag time analysis	260	257	517
With clinical progression date	259	252	511
With CA-125 doubling date	200	189	389
With both clinical and CA-125 progression dates	199	184	383

Table 3. No. of Patients With Clinical or CA-125 Date of Progression

Indicator	Clinical and Radiologic Progression	No Clinical or Radiologic Progression	Total
CA-125 doubling	383	6	389
No CA-125 doubling	128	111	239
No CA-125 assessable	45	7	52
Total	556	124	680

PFS

Figure 1A shows that the median PFS according to standard criteria was 11.5 months in the CP arm and 15.7 months in the TP arm ($P < .001$). CA-125 is used to define progression in Figure 1B and shows a similar significant trend in favor of TP, with the median progression-free and alive rates of 12.9 months in the CP arm and 16.5 months in the TP arm ($P < .0046$). In Figure 1C, the date of progression was the earliest date of progression defined by either criteria, which shortened the median PFS to 10.2 months in the CP arm and 13.8 months in the TP arm ($P < .001$), but again showed TP to be superior. However, the reduction in the median PFS observed by using the composite end point of CA-125 or clinical progression rather than clinical progression alone was not significant in either of the two trial arms. The 95% CIs for the medians overlapped in the CP arm (standard criteria: 95% CI, 10.3 to 13.2 months; earliest by either criteria: 95% CI, 9.5 to 11.6 months) as well as in the TP arm (standard criteria: 95% CI, 13.9 to 18.5 months; earliest by either criteria: 95% CI, 12.6 to 15.6 months).

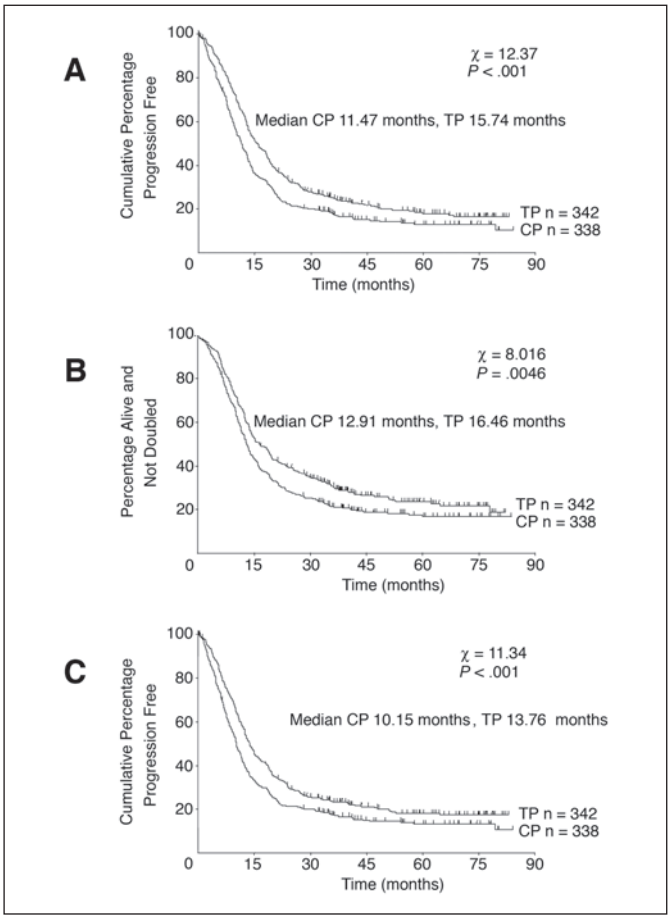

Fig 1. (A) Progression-free survival curves comparing cyclophosphamide and cisplatin (CP) with paclitaxel and cisplatin (TP), using clinical and radiologic criteria only. (B) CA-125 progression-free and alive rate comparing CP with TP (includes patients not assessable by CA-125). (C) Progression-free survival comparing CP with TP, using first date of progression defined by either clinical, radiologic, or CA-125 criteria.

Hazard Analysis

Hazard analysis using either standard criteria or CA-125 to define progression confirmed a lower overall hazard for those in the TP arm compared with those in the CP arm (Fig 2; Table 4). The hazard for both arms changed over time, with an initial increase to a peak followed by a decline. The apparent terminal increase in hazard shown in Figure 2 is an artefact; the final time point extends to infinity.

For clinical progression, the hazard difference between the arms declined significantly over time: although the overall interaction between trial arm and time interval blocks was not significant ($G = 9.699$; $P = .36$ with 9 *df*), examination of the parameter estimates revealed a trend for the hazard of the two arms to converge over time. Replacing the overall interaction with a simple linear trend for converging hazard was acceptable ($G = 2.155$; $P = .9759$ with 8 *df*) and this term was significant (Table 4). A similar analysis using CA-125 progression showed a similar trend for the hazard of the two arms to converge with time, but this was not significant (Table 4). The lack of significance probably was due to the lower number of events in this analysis, which results in lower statistical power: only 62% of patients showed CA-125 doubling (n = 389), whereas 81% showed clinical progression (n = 511).

The extent of the difference (ie, hazard ratio) between the two trial arms was also similar when comparing progression being measured by either method (Fig 3), with clinical and CA-125 hazard ratio estimates significantly correlated over the time intervals (Kendall's $\tau = 0.60$; $P = .016$). It is clear from Figure 3 that the use of CA-125 to define progression does not result in earlier detection of the difference between the arms compared with use of just the standard criteria.

Lead and Lag Time of CA-125–Defined PD

The total number of patients assessable for their lead or lag times was 517. In 61 patients, because of missing samples, the last CA-125

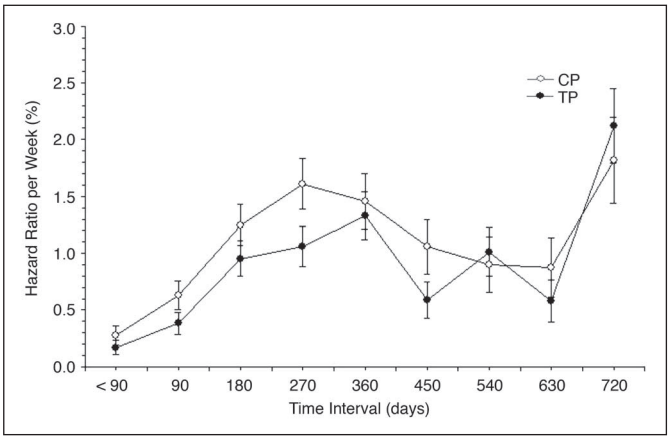

Fig 2. Hazard of CA-125 doubling during 90-day periods comparing cyclophosphamide and cisplatin (CP) with paclitaxel and cisplatin (TP).

Table 4. Significance of Results in Hazard Analysis

Model Term	Indicator of Progression	G^*	*df*	*P*
Difference between trial arms	Clinical progression	9.550	1	.0020
	CA-125 doubling	6.461	1	.0110
	Combined	7.066	1	.0079
Difference between time intervals	Clinical progression	132.3	9	< .0001
	CA-125 doubling	144.8	9	< .0001
	Combined	115.0	9	< .0001
Decreasing arm difference with time	Clinical progression	7.544	1	.0060
	CA-125 doubling	3.286	1	.0699
	Combined	6.943	1	.0084

*Change in variance.

measurement recorded was before the date of clinical progression, and these patients were considered nonassessable for the lead/lag time analysis. Among the remaining 456 patients, six experienced a CA-125 progression without a subsequent clinical progression.

There were 254 patients with time to progression (TTP) according to CA-125 ≤ TTP according to standard criteria, thus showing a lead time for CA-125 relative to clinical progression. The mean lead time in this group was 95 days (95% CI, 77 to 112 days). The date of CA-125 progression was up to 650 days before clinical progression and the median lead time was 55 days. One hundred eighty-nine patients (74%) had CA-125 progression more than 15 days before clinical progression. Figure 4A shows the lead time distribution in the group experiencing CA-125 progression before clinical progression.

There were 196 patients who had clinical progression and no CA-125 progression or clinical progression before CA-125 progression. Patients who did not experience CA-125 progression were censored at the time of the last available CA-125 value. Figure 4B shows the Kaplan-Meier estimate of the proportion of patients with CA-125 progression as a function of the time after clinical progression. In 95 patients, although the CA-125 doubling was not confirmed by a second sample, we accepted them as experiencing CA-125 progression. A confirmatory CA-125 sample was not stipulated in the trial protocol, so we could not demand it in a retrospective analysis.

The mean lag time was estimated as the area under the Kaplan-Meier TTP curve truncated at 1 year. This lag time was estimated at 173 ± 10 days. Weighting the mean lead and lag times with the number of patients in the two groups gives a net lag time of 22 days (95% CL = 4, 40 days) in TTP using CA-125 progression as the only criterion.

If progression was defined as the earliest time of CA-125 and clinical progression, there obviously would not be any patients with false-negative predictions, and the lag time in patients with TTP according to CA-125 more than TTP according to standard criteria would be zero. In this case, the mean lead time achieved by the composite end point relative to clinical progression alone is estimated at 54 days (95% CI, 44 to 64 days).

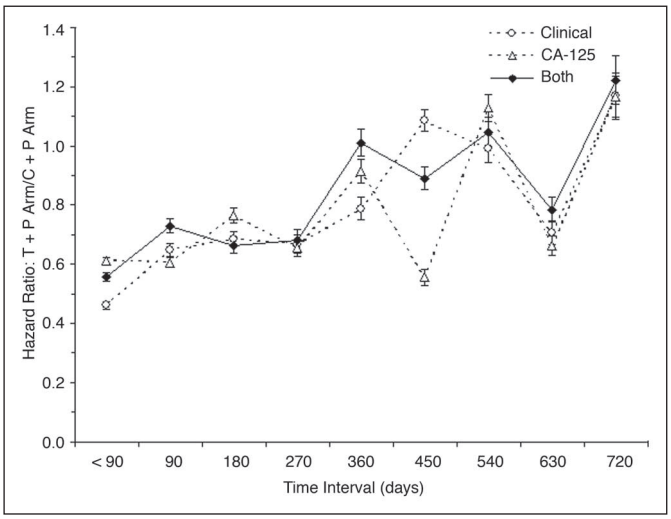

Fig 3. Hazard ratio for progression comparing use of clinical and radiologic criteria, CA-125 criteria, and both criteria. CP, cyclophosphamide and cisplatin; TP, paclitaxel and cisplatin.

DISCUSSION

These results show that the main conclusions of this trial would have been unaltered if clinical progression dates had been replaced by CA-125 doubling dates. There was a significantly reduced hazard for progression for those receiving paclitaxel and cisplatin compared with those receiving cyclophosphamide and cisplatin. In addition, the pattern of the hazard over time and the hazard ratio of the two trial arms were also similar, regardless of the method used to assess progression. Our analyses therefore suggest that the criteria for progression according to doubling of CA-125 provide an accurate and useful method of assessing progression.

Analyses from this study do not suggest that differences between trial arms could have been detected more quickly by using CA-125 doubling than by using clinical progression. CA-125 doubling occurred before clinical progression in 254 patients, but in another 196 patients with at least one CA-125 measurement after the date of clinical progression, CA-125 progression occurred later. The actuarial Kaplan-Meier estimate was used to correct for censoring; that is, to allow for patients who had not experienced CA-125 progression at the time of the last valid CA-125 measurement. Given that the false-negative rate obviously depends on the length of follow-up, it was estimated that 1 year after clinical progression was established, 12% of all assessable patients had not developed a CA-125 progression. Some patients with a lag between clinical and CA-125 progression were

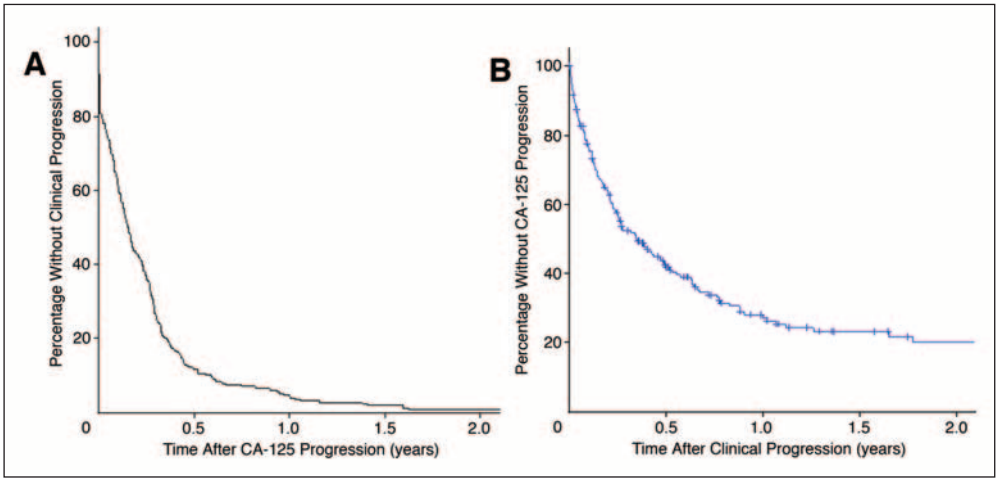

Fig 4. (A) Lead time between CA-125–defined progression and progression defined by clinical and radiologic criteria. (B) Lag time between progression defined by clinical and radiologic criteria and CA-125–defined progression.

censored, but it is reasonable to assume that a high diagnostic intensity would have helped overcome this problem. Thus, the rate of occurrence of events (progressions) may be lower when using CA-125 doubling alone as the indicator of progression, which would reduce statistical power for tests of differences between trial arms in hazard or PFS. Therefore, the sample size would have to be increased by approximately 12% to compensate for this effect. However, the use of both CA-125 and standard criteria would result in the observation of more events earlier. In this case, the mean lead time was estimated at 54 days.

It is worth noting that the trial protocol was not specifically designed with the aim of testing the adequacy of the CA-125 doubling criteria for progression. As a result, marker levels in some patients were not measured during follow-up and, in a proportion of patients, were not measured until clinical progression. We suspect that in many patients, the true dates of CA-125 progression may have been earlier than documented if appropriately dated samples had been taken. This would also have an effect on the analysis of lead and lag times. In a prospective trial with more frequent CA-125 measurements, it is likely that the proportion of patients with a CA-125 lead time would increase. When progression was defined as the earliest date of progression by either method, the median PFS was shortened (Fig 1C), although this was not significant. This is important to consider when comparisons are made between studies in which CA-125 is or is not used to define progression.

Hazard ratios are accepted as a measure of treatment effect. We believe that estimating hazard ratios as a function of time is a novel and potentially useful method for comparing different end points.

There is considerable debate about how a patient should be managed if the only evidence of progression is an elevated CA-125 level. This issue will only be resolved when the Medical Research Council and European Organisation for Research and Treatment of Cancer have reported the results of the OVO5/55955 trial. This trial is comparing early versus late treatment of relapse detected by CA-125. If one is tempted to initiate additional treatment just because of an elevated CA-125 level, it is essential to determine how accurately the CA-125 predicts progression. The specificity of CA-125 in this study was only 95%, which was considerably lower than that reported by Rustin et al,[3,10] However, of the six false-positive patients, three had only a short follow-up time after the elevated CA-125 level; in two patients, there was only a single (probably rogue) sample that predicted relapse. The requirement of a confirmatory elevated sample would almost certainly have increased the specificity to 99%. Now that there is an established CA-125 definition for progression, it would be advisable that before therapy is started to treat relapse just because of an increasing CA-125, clinicians wait until the levels satisfy the GCIG definition for progression.

Many investigators perform radiographic assessments when CA-125 levels increase; if the assessment shows disease progression, they use that as their progression end point. If the radiographic assessments are negative but the patient has symptoms compatible with relapse, most would instigate therapy; in this example, we would suggest that the CA-125 date of progression be used as an end point. If the radiographic assessments are negative and the patient has no symptoms, some patients will want immediate therapy and some will be prepared to wait until they develop symptoms or until signs of progression develop. To prevent bias in declaring the date of progression, it is important to perform investigations such as CA-125 measurements and scans at predetermined times on all arms of randomized trials, unless they are prompted by symptoms suggesting progression. Once a patient and his or her physician are aware of increasing CA-125 levels, it becomes difficult to dictate from a protocol the timing of additional radiographic assessments just so that a RECIST end point can be declared. Pragmatism suggests the value of CA-125 as an end point in this situation; however, until CA-125–defined progression has been validated fully, trials should continue to be analyzed using the different criteria separately. Unless enough patients have end points defined by RECIST, inappropriate conclusions about efficacy could occur, and it will not be possible to validate the CA-125 criteria.

We recommend that data be collected prospectively in current and future trials so that PFS can be determined according to CA-125 in addition to RECIST criteria. Regulatory authorities will require data from more trials supporting our validation before they accept a new end point. Although this study only investigated the first relapse, there is no reason to believe that the CA-125 definition would not be just as reliable for defining progression after later treatments. It is essential that CA-125 measurements be performed at regular, predetermined intervals, preferably every 2 to 3

months during follow-up. The use of the recently proposed GCIG definition of response,[16] plus the progression definition, will lead to faster accrual of patients into ovarian cancer trials and more events. This will result in the more rapid assessment of new and existing therapies, and it is hoped, earlier approval by the regulatory authorities.

REFERENCES

1. Van der Burg MEL, Lammes FB, Verweij J: The role of CA 125 in the early diagnosis of progressive disease in ovarian cancer. Ann Oncol 1:301-302, 1990

2. Tuxen MK, Soletormos G, Dombernowsky P: Tumor markers in the management of patients with ovarian cancer. Cancer Treat Rev 21:215-245, 1995

3. Rustin GJ, Nelstrop AE, Tuxen MK, et al: Defining progression of ovarian carcinoma during follow-up according to CA 125: A North Thames Ovary Group Study. Ann Oncol 7:361-364, 1996

4. Muggia FM, Braly PS, Brady MF, et al: Phase III randomized study of cisplatin versus paclitaxel versus cisplatin and paclitaxel in patients with suboptimal stage III or IV ovarian cancer: A Gynecologic Oncology Group study. J Clin Oncol 18:106-115, 2000

5. Vergote I, Rustin GJS, Eisenhauer EA, et al: Re: New guidelines to evaluate the response to treatment in solid tumours (ovarian cancer). J Natl Cancer Inst 92:1534-1535, 2000

6. Bast RC, Klug TL, John ES: A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med 309:883-887, 1983

7. van der Burg ME, Lammes FB, Verweij J: The role of CA 125 in the early diagnosis of progressive disease in ovarian cancer. Ann Oncol 1:301-302, 1990

8. Krebs HB, Goplerud DR, Kilpatrick SJ, et al: The role of CA 125 as a tumour marker in ovarian cancer. Obstet Gynecol 67:473-477, 1986

9. Rustin GJ, Nelstrop A, Stilwell J, et al: Savings obtained by CA-125 measurements during therapy for ovarian carcinoma: The North Thames Ovary Group. Eur J Cancer 28:79-82, 1992

10. Rustin GJS, Marples M, Nelstrop AE, et al: Use of CA 125 to define progression of ovarian cancer in patients with persistently elevated levels. J Clin Oncol 19:4054-4057, 2001

11. Therasse P, Arbuck SG, Eisenhauer EA, et al: New guidelines to evaluate the response to treatment in solid tumours. J Natl Cancer Inst 92:205-216, 2000

12. Piccart MJ, Bertelsen K, James K, et al: Randomized intergroup trial of cisplatin-paclitaxel versus cisplatin-cyclophosphamide in women with advanced epithelial ovarian cancer: Three-year results. J Natl Cancer Inst 92:699-708, 2000

13. Kaplan EI, Meier P: Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457-481, 1958

14. Anderson JR, Crowley JJ, Propert KJ: Interpretation of survival data in clinical trials (with discussion by R. Simon). Oncology 5:104-110, 1991

15. Aitkin M, Anderson D, Francis B, et al: Statistical Modelling in GLIM. Oxford, United Kingdom, Clarendon Press, 1989

16. Rustin GJS, Quinn M, Thigpen T, et al: New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). J Natl Cancer Inst 96:487-488, 2004

Acknowledgment

We thank all of the investigators who contributed patients and data for this analysis and the data centers of the four organizations for their work in collecting the data.

Authors' Disclosures of Potential Conflicts of Interest

Although all authors completed the disclosure declaration, the following authors or their immediate family members indicated a financial interest. No conflict exists for drugs or devices used in a study if they are not being evaluated as part of the investigation. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Authors	Employment	Leadership	Consultant	Stock	Honoraria	Research Funds	Testimony	Other
Martine J. Piccart			Aventis (A); BMS (A); GSK (A)		AstraZeneca (A); Aventis (A); Novartis (A)			
James Cassidy			Roche (A); Sanofi (A); Aventis (A)		Roche (A); Sanofi (A); Pfizer (A); Roche (B)			

Dollar Amount Codes (A) < $10,000 (B) $10,000-99,999 (C) ≥ $100,000 (N/R) Not Required

Author Contributions

Conception and design: Gordon J.S. Rustin, Soeren M. Bentzen, Martine J. Piccart, Kamma Bertelsen, Elizabeth Eisenhauer
Financial support: Gordon J.S. Rustin
Administrative support: Gordon J.S. Rustin, Gavin Shreeves, Benoit Baron, Gavin Stuart
Provision of study materials or patients: James Cassidy
Collection and assembly of data: Petra Timmers, Ann Nelstrop, Gavin Shreeves, Benoit Baron, Kamma Bertelsen, Elizabeth Eisenhauer
Data analysis and interpretation: Gordon J.S. Rustin, Petra Timmers, Ann Nelstrop, Gavin Shreeves, Soeren M. Bentzen, Benoit Baron, Gavin Stuart
Manuscript writing: Gordon J.S. Rustin, Petra Timmers, Gavin Shreeves, Soeren M. Bentzen, Benoit Baron, Gavin Stuart
Final approval of manuscript: Gordon J.S. Rustin, Ann Nelstrop, Soeren M. Bentzen, Benoit Baron, Martine J. Piccart, Gavin Stuart